GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 13, 2001, 12:22:04 ; Search time 59.73 Seconds
(without alignments)
534.888 Million cell updates/sec

Title: US-09-784-340-2
Perfect score: 2802
Sequence: 1 MRSDKSALVFLLLQLFCVGC.....KCFLFSCQKFNKTRKIEKRE 527

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_0601:*
1: /SIDS1/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1924.5	68.7	529	21	AAB28677	Human carbohydrate
2	1732	61.8	530	19	AAW47126	Uridine diphospho-
3	1729	61.7	528	21	AAY78933	Human UDP-glucuron
4	1713	61.1	530	21	AAY78935	Human UDP-glucuron
5	1673.5	59.7	524	21	AAY78934	Human UDP-glucuron
6	1120	40.0	533	13	AAR26153	HUG-Br1. Homo sap
7	1058.5	37.8	534	13	AAR26154	HUG-Br2. Homo sap
8	755.5	27.0	245	21	AAY57100	UDP-glucuronosyltr
9	714.5	25.5	523	21	AAB24025	Human PRO1780 prot
10	714.5	25.5	523	21	AAY99419	Human PRO1780 (UNQ
11	714.5	25.5	523	22	AAB66168	Protein of the inv
12	708.5	25.3	523	22	AAB88348	Human membrane or
13	405	14.5	78	21	AAG03280	Human secreted pro
14	391	14.0	129	20	AAY29525	Human lung tumour
15	391	14.0	129	21	AAB44411	Human lung tumour-
16	383.5	13.7	288	21	AAY57092	UDP-glucuronosyltr
17	349.5	12.5	310	21	AAY57098	UDP-glucuronosyltr
18	341	12.2	287	13	AAR30189	UGT1F Exon 1 produ
19	341	12.2	287	21	AAY57096	UDP-glucuronosyltr
20	327.5	11.7	317	21	AAY57097	UDP-glucuronosyltr
21	326	11.6	289	13	AAR30190	UGT1E Exon 1 produ
22	326	11.6	289	21	AAY57093	UDP-glucuronosyltr
23	321	11.5	289	21	AAY57094	UDP-glucuronosyltr
24	320	11.4	289	21	AAY57095	UDP-glucuronosyltr
25	317.5	11.3	253	21	AAY57099	UDP-glucuronosyltr
26	317	11.3	516	20	AAY08462	M. brassicae ecdys
27	316	11.3	289	13	AAR30192	UGT1C Exon 1 produ
28	297.5	10.6	98	13	AAR30166	UGT1 Exon 5 produc
29	294.5	10.5	248	13	AAR30191	UGT1D Exon 1 produ
30	274.5	9.8	515	19	AAW56750	Ecdysteroid UDP-gl
31	262	9.4	243	13	AAR30194	UGT1A Exon 1 produ
32	260	9.3	94	21	AAB53721	Human colon cancer
33	254.5	9.1	506	12	AAR10429	Ecdysteroid UDP-gl
34	235	8.4	74	13	AAR30165	UGT1 Exon 4 produc
35	190	6.8	466	18	AAW09825	UDP-glucose:thiohy
36	175	6.2	399	12	AAR13989	Zeaxanthin glycosy
37	175	6.2	431	11	AAR07464	Polypeptide with e
38	175	6.2	431	20	AAW87890	Protein encoded by
39	171.5	6.1	68	21	AAB56504	Human prostate can
40	171.5	6.1	439	21	AAG04893	Arabidopsis thalia
41	171.5	6.1	445	21	AAG04892	Arabidopsis thalia
42	171.5	6.1	460	21	AAG40011	Arabidopsis thalia
43	171.5	6.1	465	21	AAG40010	Arabidopsis thalia
44	171.5	6.1	480	21	AAG04891	Arabidopsis thalia
45	171.5	6.1	481	21	AAG40009	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAB28677
ID AAB28677 standard; Protein; 529 AA.
XX
AC AAB28677;
XX
DT 13-FEB-2001 (first entry)
XX
DE Human carbohydrate-modifying enzyme Incyte ID No: 2912330CD1.
XX
KW Human; carbohydrate-modifying enzyme; CME; antidiabetic;
KW immunosuppressive; anti-HIV; antiinflammatory; antianaemic;
KW antiasthmatic; antiarteriosclerotic; antithyroid; hepatotropic;
KW nephrotropic; antigout; thyromimetic; neuroprotective; osteopathic;
KW antiarthritic; antipsoriatic; uropathic; ophthalmological;
KW dermatological; antiulcer; cytostatic; virucide; antibacterial;
KW fungicide; protozoacide; tranquiliser; vulnerary; diabetes;
KW autoimmune disorder; inflammatory disorder; infection.
XX
OS Homo sapiens.
XX
PN WO200063351-A2.
XX
PD 26-OCT-2000.
XX
PF 20-APR-2000; 2000WO-US10882.
XX
PR 21-APR-1999; 99US-0130383.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Lal P, Yue H, Tang YT, Hillman JL, Baughn MR, Yang J;
XX
DR WPI; 2000-672729/65.

```
DR   N-PSDB; AAC65396.
XX
PT   Novel carbohydrate modifying enzyme polypeptides and polynucleotides
PT   for diagnosis, treatment, and prevention of carbohydrate metabolism
PT   disorders, autoimmune/inflammatory disorders, and cancer -
XX
PS   Claim 1; Page 71-72; 75pp; English.
XX
CC   The present sequence is a human carbohydrate-modifying enzyme
CC   (CME). CME polynucleotides and polypeptides are useful for treating and
CC   diagnosing diseases associated with CME such as diabetes,
CC   autoimmune/inflammatory disorders such as AIDS, Addison's disease,
CC   adult respiratory distress syndrome, allergies, anaemia, asthma,
CC   atherosclerosis, autoimmune thyroiditis, bronchitis, cholocystitis,
CC   contact dermatitis, Crohn's disease, emphysema, erythroblastosis fetalis,
CC   glomerulonephritis, Good pasture's syndrome, gout, Grave's disease,
CC   Hashimoto's thyroiditis, multiple sclerosis, myasthenia gravis,
CC   osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis,
CC   Reiter's syndrome, arthritis, scleroderma, sjogren's syndrome, systemic
CC   lupus erythematosus, ulcerative colitis, uveitis, Werner syndrome,
CC   complications of cancer, haemodialysis, and extracorporeal circulation,
CC   viral, bacterial, fungal parasitic, protozoal, and helminthic infections,
CC   trauma, or cancer. CME, or its catalytic or immunogenic fragment, is
CC   useful for drug screening.
XX
SQ   Sequence   529 AA;

  Query Match             68.7%;  Score 1924.5;  DB 21;  Length 529;
  Best Local Similarity   69.9%;  Pred. No. 1.5e-197;
  Matches  369;  Conservative   47;  Mismatches  109;  Indels    3;  Gaps    2;

Qy       3 SDKSALVFLLLQLFC-VGCGFCGKVLVWPCDMSHWLNVKVILEELIVRGHEVTVLTHSKP 61
Db       2 smkwtsallliqlscyfssgscgkvlvwptefshwmniktildelvqrghevtvlassas 61

Qy      62 SLIDYRKPSALKFEVVHMPQDRTEENEIFVDLALN--VLPGLSTWQSVIKLNDFFVEIRG 119
Db      62 isfdpnspstlkfevypvsltktefediikqlvkrwaelpkdtfwsyfsqvqeimwtfnd 121

Qy     120 TLKMMCESFIYNQTLMKKLQETNYDVMLIDPVIPCGDLMAELLAVPFVLTLRISVGGNME 179
Db     122 ilrkfckdivsnkklmkklqesrfdvvladavfpfgellaellkipfvyslrfspgyaie 181

Qy     180 RSCGKLPAPLSYVPVPMTGLTDRMTFLERVKNSMLSVLFHFWIQDYDYHFWEEFYSKALG 239
Db     182 khsggllfppsyvpvvmselsdqmtfiervknmiyvlyfefwfqifdmkkwdqfysevlg 241

Qy     240 RPTTLCETVGKAEIWLIRTYWDFEFPQPYQPNFEFVGGLHCKPAKALPKEMENFVQSSGE 299
Db     242 rpttlsetmakadiwlirnywdfqfphpllpnvefvgglhckpakplpkemeefvqssge 301

Qy     300 DGIVVFSLGSLFQNVTEEKANIIASALAQIPQKVLWRYKGKKPSTLGANTRLYDWIPQND 359
Db     302 ngvvvfslgsmvsntseeranviasalakipqkvlwrfdgnkpdtlglntrlykwipqnd 361

Qy     360 LLGHPKTKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKTMT 419
Db     362 llghpktkafithggmngiyeaiyhgvpmvgvpifgdqldniahmkakgaaveinfktmt 421

Qy     420 SEDLLRALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSAAH 479
Db     422 sedllralrtvitdssykenamrlsrihhdqpvkpldravfwiefvmrhkgakhlrsaah 481

Qy     480 DLTWFQHYSIDVIGFLLTCVATAIFLFTKCFLFSCQKFNKTRKIEKRE 527
Db     482 dltwfqhysidvigflltcvataiflftkcflfscqkfnktrkiekre 529


RESULT   2
AAW47126
ID   AAW47126 standard; Protein; 530 AA.
XX
AC   AAW47126;
XX
DT   26-MAY-1998  (first entry)
XX
DE   Uridine diphospho-glucuronosyltransferase 2B17 (UGT2B17) enzyme.
XX
KW   Uridine diphospho-glucuronosyltransferase 2B17; UGT2B17; catalyse;
KW   androsterone; androsterone-glucuronic acid; androgen; enzyme.
XX
OS   Homo sapiens.
XX
PN   WO9744466-A1.
XX
PD   27-NOV-1997.
XX
PF   16-MAY-1997;   97WO-CA00328.
XX
PR   17-MAY-1996;   96US-0649319.
XX
PA   (ENDO-) ENDORECHERCHE INC.
XX
PI   Beaulieu M,  Belanger A,  Hum DW,  Levesque E;
XX
DR   WPI; 1998-018520/02.
DR   N-PSDB; AAV15900.
XX
PT   DNA encoding uridine di:phospho:glucuronosyl:transferase 2B17 -
PT   which catalyses conversion of androsterone to
PT   androsterone-glucuronic acid
XX
PS   Claim 16; Pages 4-6; 53pp; English.
XX
CC   This is the enzyme uridine diphospho-glucuronosyltransferase 2B17
CC   (UGT2B17). This novel enzyme catalyses the conversion of androsterone
CC   to androsterone-glucuronic acid. The UGT2B17 can be used to detect
CC   anti-UGT2B17 antibodies. The antibody can be used to detect a localised
CC   concentration of UGT2B17 or an alteration in androgen activity. The
CC   UGT2B17 can also be used to alter the concentration of an androgenic
CC   compound in a tissue, specifically dihydrotestosterone. An isolated
CC   nucleotide sequence comprising at least 30 consecutive nucleotides from
CC   the coding region of the 2107 base pair sequence, or its complement can
CC   be used to block the synthesis of UGT2B17, e.g. an expression disrupting
CC   sense or antisense fragment, or as a probe for a UGT2B17 coding sequence.
XX
SQ   Sequence   530 AA;

  Query Match             61.8%;  Score 1732;  DB 19;  Length 530;
  Best Local Similarity   61.5%;  Pred. No. 7.1e-177;
  Matches  326;  Conservative   74;  Mismatches  112;  Indels   18;  Gaps    3;

Qy       9 VFLLLQLFC-VGCGFCGKVLVWPCDMSHWLNVKVILEELIVRGHEVTVLTHSKPSLIDYR 67
Db       8 vfllmqlscyfssgscgkvlvwpteyshwinmktileelvqrghevivltssasilvnas 67

Qy      68 KPSALKFEVVHMPQDRTEENEIFVDLALNVLPGLSTWQSVIKLNDFFVEIRGTLKM---- 123
Db      68 kssaiklevyptsltkndledffmkm-------fdrwtysisknfwsyfsqlqelcwey 120

Qy     124 ------MCESFIYNQTLMKKLQETNYDVMLIDPVIPCGDLMAELLAVPFVLTLRISVGGN 177
Db     121 sdyniklcedavlnkklmrklqeskfdvlladavnpcgellaellnipflyslrfsvgyt 180

Qy     178 MERSCGKLPAPLSYVPVPMTGLTDRMTFLERVKNSMLSVLFHFWIQDYDYHFWEEFYSKA 237
Db     181 vekngggflfppsyvpvvmselsdqmifmerikmmiymlyfdfwfqaydlkkwdqfysev 240

Qy     238 LGRPTTLCETVGKAEIWLIRTYWDFEFPQPYQPNFEFVGGLHCKPAKALPKEMENFVQSS 297
Db     241 lgrpttlfetmgkaemwlirtywdfefprpflpnvdfvgglhckpakplpkemeefvqss 300

Qy     298 GEDGIVVFSLGSLFQNVTEEKANIIASALAQIPQKVLWRYKGKKPSTLGANTRLYDWIPQ 357
```

```
Db    301 gengivvfslgsmisnmseesanmiasalaqipqkvlwrfdgkkpntlgsntrlykwlpq 360

Qy    358 NDLLGHPKTKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKT 417
          |||||||||||||||| |||||||||||:||||:|:| || ||||||||||||: :: :|
Db    361 ndllghpktkafithggtngiyeaiyhgipmvgiplfadqhdniahmkakgaalsvdirt 420

Qy    418 MTSEDLLRALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSA 477
          |:| ||| ||::|| |  |||| |:|||||||||||||||||||||||||||||||| |
Db    421 mssrdllnalksvindpiykenimklsrihhdqpvkpldravfwiefvmrhkgakhlrva 480

Qy    478 AHDLTWFQHYSIDVIGFLLTCVATAIFLFTKCFLFSCQKFNKTRKIEKRE 527
          ||:||| |::|:||| ||| |||| ||: ||| ||  :|  || | :||:
Db    481 ahnltwiqyhsldviafllacvatmifmitkcclfcfrklaktgkkkkrd 530


RESULT   3
AAY78933
ID   AAY78933 standard; Protein; 528 AA.
XX
AC   AAY78933;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Human UDP-glucuronosyltransferase 2B4 amino acid sequence.
XX
KW   UDP-glucuronosyltransferase 2B4; UGT2B4; polymorphism; metabolism; SNPs;
KW   drug interaction; detect; human; single nucleotide polymorphism.
XX
OS   Homo sapiens.
XX
PN   WO200006776-A1.
XX
PD   10-FEB-2000.
XX
PF   22-JUL-1999;   99WO-US16675.
XX
PR   28-JUL-1998;   98US-0094391.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   Galvin M,  Miller A,  Penny L,  Riedy M;
XX
DR   WPI; 2000-195321/17.
DR   N-PSDB; AAZ95199.
XX
PT   Novel human UDP-glucuronosyltransferase sequence, polymorphisms for
PT   genotyping individuals to predict rate of metabolism of substrates and
PT   for identifying potential drug interactions  -
XX
PS   Disclosure; Page 36-37; 72pp; English.
XX
CC   This sequence represents the human UDP-glucuronosyltransferase 2B4
CC   (UGT2B4) amino acid sequence. UDP-glucuronosyltransferase (UGTs) are a
CC   family of enzymes that catalyse the glucuronic acid conjugation of a
CC   wide range of endogenous and exogenous substrates. The UGT2B gene
CC   subfamily encode steroid metabolizing isoforms in the liver. Alteration
CC   of the expression or function of UGTs may effect drug metabolism. The
CC   invention relates to non-chromosomal nucleic acid molecules, which
CC   comprise human UGT2B sequence polymorphisms (see AAZ95051-Z95110). Probes
CC   which detect the UGT2B locus polymorphisms can be used to detect altered
CC   UGT2B metabolism of a substrate in an individual. The nucleic acid
CC   molecules comprising a human UGT2B sequence polymorphism can be used in
CC   screening assays for genotyping individuals, also to predict their rate
CC   of metabolism of UGT2B substrate, potential drug-drug interactions and
CC   adverse side effects. The polymorphisms can be used as single nucleotide
CC   polymorphisms (SNPs) for detecting genetic linkage related to phenotypic
CC   variation in activity or expression of UGT2B protein. The polymorphism
CC   containing nucleic acid molecules may also be used for generating
CC   genetically modified non-human animals and for obtaining site specific
CC   gene modification in cell lines.
XX
SQ   Sequence   528 AA;
```

```
  Query Match             61.7%;  Score 1729;  DB 21;  Length 528;
  Best Local Similarity   62.9%;  Pred. No. 1.5e-176;
  Matches  332;  Conservative   61;  Mismatches  131;  Indels    4;  Gaps    3;

Qy      3 SDKSALVFLLLQLFC-VGCGFCGKVLVWPCDMSHWLNVKVILEELIVRGHEVTVLTHSKP 61
            | |     ||:|| |    | |||||||| : |||:|:| ||:||: ||||||||  |
Db      2 smkwtsallliqlscyfssgscgkvlvwptefshwmniktildelvqrghevtvlassas 61

Qy     62 SLIDYRKPSALKFEVVHMPQDRTEENEIFVDLALN--VLPGLSTWQSVIKLNDFFVEIRG 119
              |   || |||||  :   :||  :|   |      ||  : |    :: :
Db     62 isfdpnspstlkfevypvsltktefediikqlvkrwaelpkdtfwsyfsqvqeimwtfnd 121

Qy    120 TLKMMCESFIYNQTLMKKLQETNYDVMLIDPVIPCGDLMAELLAVPFVLTLRISVGGNME 179
           |:  |:  : |: |||||||: :||:| | | | |:|:|||| :||| :|| | |  :|
Db    122 ilrkfckdivsnkklmkklqesrfdvvladavfpfgellaellkipfvyslrfspgyaie 181

Qy    180 RSCGKLPAPLSYVPVPMTGLTDRMTFLERVKNSMLSVLFHFWIQDYDYHFWEEFYSKALG 239
          :  | |  | ||||| |: |:|:|||:||||| :  : | || | :|   |::|||: ||
Db    182 khsggllfppsyvpvvmselsdqmtfiervknmiyvlyfefwfqifdmkkwdqfysevlg 241

Qy    240 RPTTLCETVGKAEIWLIRTYWDFEFPQPYQPNFEFVGGLHCKPAKALPKEMENFVQSSGE 299
          ||||| ||: ||:|||| |||:|| |  || |||||||||||| |||||| |||||||
Db    242 rpttlsetmakadiwlirnywdfqfphpllpnvefvgglhckpakplpkemeefvqssge 301

Qy    300 DGIVVFSLGSLFQNVTEEKANIIASALAQIPQKVLWRYKGKKPSTLGANTRLYDWIPQND 359
          :|:|||||||:  | :||:||:|||||:|||||||: | || ||| ||||| ||||||
Db    302 ngvvvfslgsmvsntseeranviasalakipqkvlwrfdgnkpdtlglntrlykwipqnd 361

Qy    360 LLGHPKTKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKTMT 419
          |||||||:|||||||| ||||||||||:||||||:| || |||||||||||| ::| ||:
Db    362 llghpktrafithggangiyeaiyhgipmvgvplfadqpdniahmkakgaavsldfhtms 421

Qy    420 SEDLLRALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSAAH 479
          | ||| ||:||| |  ||||||:|||||||||||||||||||||||||||||||| |||
Db    422 stdllnalktvindplykenamklsrihhdqpvkpldravfwiefvmrhkgakhlrvaah 481

Qy    480 DLTWFQHYSIDVIGFLLTCVATAIFLFTKCFLFSCQKFNKTRKIEKRE 527
          ||||||::|:|| |||| |||| ||: ||| ||   || :| |  ||:
Db    482 dltwfqyhsldvtgfllacvatvifiitkc-lfcvwkfvrtgkkgkrd 528


RESULT   4
AAY78935
ID   AAY78935 standard; Protein; 530 AA.
XX
AC   AAY78935;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Human UDP-glucuronosyltransferase 2B15 amino acid sequence.
XX
KW   UDP-glucuronosyltransferase 2B15; UGT2B15; polymorphism; metabolism;
KW   drug interaction; detect; human; single nucleotide polymorphism; SNPs.
XX
OS   Homo sapiens.
XX
PN   WO200006776-A1.
XX
PD   10-FEB-2000.
XX
PF   22-JUL-1999;   99WO-US16675.
XX
PR   28-JUL-1998;   98US-0094391.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   Galvin M,  Miller A,  Penny L,  Riedy M;
XX
DR   WPI; 2000-195321/17.
DR   N-PSDB; AAZ95206.
XX
PT   Novel human UDP-glucuronosyltransferase sequence, polymorphisms for
```

PT genotyping individuals to predict rate of metabolism of substrates and
PT for identifying potential drug interactions -
XX
PS Disclosure; Page 59-60; 72pp; English.
XX
CC This sequence represents the human UDP-glucuronosyltransferase 2B15
CC (UGT2B15) amino acid sequence. UDP-glucuronosyltransferase (UGTs) are a
CC family of enzymes that catalyse the glucuronic acid conjugation of a
CC wide range of endogenous and exogenous substrates. The UGT2B gene
CC subfamily encode steroid metabolizing isoforms in the liver. Alteration
CC of the expression or function of UGTs may effect drug metabolism. The
CC invention relates to non-chromosomal nucleic acid molecules, which
CC comprise human UGT2B sequence polymorphisms (see AAZ95051-Z95110). Probes
CC which detect the UGT2B locus polymorphisms can be used to detect altered
CC UGT2B metabolism of a substrate in an individual. The nucleic acid
CC molecules comprising a human UGT2B sequence polymorphism can be used in
CC screening assays for genotyping individuals, also to predict their rate
CC of metabolism of UGT2B substrate, potential drug-drug interactions and
CC adverse side effects. The polymorphisms can be used as single nucleotide
CC polymorphisms (SNPs) for detecting genetic linkage related to phenotypic
CC variation in activity or expression of UGT2B protein. The polymorphism
CC containing nucleic acid molecules may also be used for generating
CC genetically modified non-human animals and for obtaining site specific
CC gene modification in cell lines.
XX
SQ Sequence 530 AA;

Query Match 61.1%; Score 1713; DB 21; Length 530;
Best Local Similarity 62.0%; Pred. No. 7.8e-175;
Matches 327; Conservative 74; Mismatches 114; Indels 12; Gaps 5;

```
Qy       9 VFLLLQLFC-VGCGFCGKVLVWPCDMSHWLNVKVILEELIVRGHEVTVLTHSKPSLIDYR 67
           ||||:|| |    | |||||||| : |||:|:| |||||: |||||||| |  :|::
Db       8 vflliqlscyfssgscgkvlvwpteyshwinmktileelvqrghevtvltssastlvnas 67

Qy      68 KPSALKFEVVHMPQDRTEENEIFVDLALNVLP----GLST---WQSVIKLNDFFVEIRGT 120
           | ||:| ||   |   |: :   |  | :|     |:|    |    :| :   |
Db      68 kssaiklev--yptsltknd--ledsllkildrwiygvsskntfwsyfsqlqelcweyydy 123

Qy     121 LKMMCESFIYNQTLMKKLQETNYDVMLIDPVIPCGDLMAELLAVPFVLTLRISVGGNMER 180
           :|:  : |: || |||| :|||:| | : |||:|:|||  :||: :|| |||   |:
Db     124 snklckdavlnkklmmklqeskfdviladalnpcgellaelfnipflyslrfsvgytfek 183

Qy     181 SCGKLPAPLSYVPVPMTGLTDRMTFLERVKNSMLSVLFHFWIQDYDYHFWEEFYSKALGR 240
           : |    | ||||| |: |:|:| |:||:|| :  : | || | ||   |::|||: |||
Db     184 ngggflfppsyvpvvmselsdqmifmeriknmihmlyfdfwfqiydlkkwdqfysevlgr 243

Qy     241 PTTLCETVGKAEIWLIRTYWDFEFPQPYQPNFEFVGGLHCKPAKALPKEMENFVQSSGED 300
           |||| ||:|||||:|||||||||||:|: || :|||||||||| |||||| |||||||:
Db     244 pttlfetmgkaemwlirtywdfefprpflpnvdfvgglhckpakplpkemeefvqssgen 303

Qy     301 GIVVFSLGSLFQNVTEEKANIIASALAQIPQKVLWRYKGKKPSTLGANTRLYDWIPQNDL 360
           |||||||||:  |::|| ||:|||||||||||||: ||||:|||:||||| |:|||||
Db     304 givvfslgsmisnmseesanmiasalaqipqkvlwrfdgkkpntlgsntrlykwlpqndl 363

Qy     361 LGHPKTKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKTMTS 420
           |||||||||||||| |||||||||:||||:|:| || ||||||||||| :: :|||:|
Db     364 lghpktkafithggtngiyeaiyhgipmvgiplfadqhdniahmkakgaalsvdirtmss 423

Qy     421 EDLLRALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSAAHD 480
           ||| ||::|| |  |||| |:|||||||||:|||||||||||||||||||||| |||:
Db     424 rdllnalksvindpvykenvmklsrihhdqpmkpldravfwiefvmrhkgakhlrvaahn 483

Qy     481 LTWFQHYSIDVIGFLLTCVATAIFLFTKCFLFSCQKFNKTRKIEKRE 527
           ||| |::|:||| ||| |||| ||: ||  ||  :|  || | :|||:
Db     484 ltwiqyhsldviafllacvatvifiitkfclfcfrklaktgkkkkrd 530
```

RESULT 5
AAY78934
ID AAY78934 standard; Protein; 524 AA.
XX

AC AAY78934;
XX
DT 05-JUN-2000 (first entry)
XX
DE Human UDP-glucuronosyltransferase 2B7 amino acid sequence.
XX
KW UDP-glucuronosyltransferase 2B4; UGT2B4; polymorphism; metabolism; SNPs;
KW drug interaction; detect; human; single nucleotide polymorphism.
XX
OS Homo sapiens.
XX
PN WO200006776-A1.
XX
PD 10-FEB-2000.
XX
PF 22-JUL-1999; 99WO-US16675.
XX
PR 28-JUL-1998; 98US-0094391.
XX
PA (AXYS-) AXYS PHARM INC.
XX
PI Galvin M, Miller A, Penny L, Riedy M;
XX
DR WPI; 2000-195321/17.
DR N-PSDB; AAZ95200.
XX
PT Novel human UDP-glucuronosyltransferase sequence, polymorphisms for
PT genotyping individuals to predict rate of metabolism of substrates and
PT for identifying potential drug interactions -
XX
PS Disclosure; Page 44-45; 72pp; English.
XX
CC This sequence represents the human UDP-glucuronosyltransferase 2B7
CC (UGT2B7) amino acid sequence. UDP-glucuronosyltransferase (UGTs) are a
CC family of enzymes that catalyse the glucuronic acid conjugation of a
CC wide range of endogenous and exogenous substrates. The UGT2B gene
CC subfamily encode steroid metabolizing isoforms in the liver. Alteration
CC of the expression or function of UGTs may effect drug metabolism. The
CC invention relates to non-chromosomal nucleic acid molecules, which
CC comprise human UGT2B sequence polymorphisms (see AAZ95051-Z95110). Probes
CC which detect the UGT2B locus polymorphisms can be used to detect altered
CC UGT2B metabolism of a substrate in an individual. The nucleic acid
CC molecules comprising a human UGT2B sequence polymorphism can be used in
CC screening assays for genotyping individuals, also to predict their rate
CC of metabolism of UGT2B substrate, potential drug-drug interactions and
CC adverse side effects. The polymorphisms can be used as single nucleotide
CC polymorphisms (SNPs) for detecting genetic linkage related to phenotypic
CC variation in activity or expression of UGT2B protein. The polymorphism
CC containing nucleic acid molecules may also be used for generating
CC genetically modified non-human animals and for obtaining site specific
CC gene modification in cell lines.
XX
SQ Sequence 524 AA;

Query Match 59.7%; Score 1673.5; DB 21; Length 524;
Best Local Similarity 61.0%; Pred. No. 1.3e-170;
Matches 319; Conservative 66; Mismatches 135; Indels 3; Gaps 3;

```
Qy       3 SDKSALVFLLLQL-FCVGCGFCGKVLVWPCDMSHWLNVKVILEELIVRGHEVTVLTHSKP 61
           | |   | ||:|| ||   | |||||||  : |||:|:| ||:||| ||||||||  |
Db       2 svkwtsvilliqlsfcfssgncgkvlvwaaeyshwmniktildeligrghevtvlassas 61

Qy      62 SLIDYRKPSALKFEVVHMPQDRTE-ENEIFVDL-ALNVLPGLSTWQSVIKLNDFFVEIRG 119
             | |    |||| |:      :|| || |   :.  : ||  : |    :: :
Db      62 ilfdpnnssalkieiyptsltktelenfimqqikrwsdlpkdtfwlyfsqvqeimsifgd 121

Qy     120 TLKMMCESFIYNQTLMKKLQETNYDVMLIDPVIPCGDLMAELLAVPFVLTLRISVGGNME 179
             :  |:  : |:  |||:||: :||:  | : || :|:|||  :||| :|  | |   |
Db     122 itrkfckdvvsnkkfmkkvqesrfdvifadaifpcsellaelfnipfvyslsfspgytfe 181

Qy     180 RSCGKLPAPLSYVPVPMTGLTDRMTFLERVKNSMLSVLFHFWIQDYDYHFWEEFYSKALG 239
           :  |    | ||||| |: |||:|||:||||| :  : | || : :|   |::|||: ||
```

```
Db    182 khsggfifppsyvpvvmseltdqmtfmervknmiyvlyfdfwfeifdmkkwdqfysevlg 241

Qy    240 RPTTLCETVGKAEIWLIRTYWDFEFPQPYQPNFEFVGGLHCKPAKALPKEMENFVQSSGE 299
          ||||| ||:|||::|||| |:|:|| |  || :|||||||||||| ||||||:|||||||
Db    242 rpttlsetmgkadvwlirnswnfqfpypllpnvdfvgglhckpakplpkemedfvqssge 301

Qy    300 DGIVVFSLGSLFQNVTEEKANIIASALAQIPQKVLWRYKGKKPSTLGANTRLYDWIPQND 359
          :|:|||||||:  |:|||:||:|||||||||||||||: | || ||| ||||| ||||||
Db    302 ngvvvfslgsmvsnmteeranviasalaqipqkvlwrfdgnkpdtlglntrlykwipqnd 361

Qy    360 LLGHPKTKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKTMT 419
          |||||||:||||||| |||||||||||:||||:|:| || |||||||||:|||| ::| ||:
Db    362 llghpktrafithggangiyeaiyhgipmvgiplfadqpdniahmkargaavrvdfntms 421

Qy    420 SEDLLRALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSAAH 479
          | ||| ||: || | ||||| |:|||| |||||||||||||||||||||||||||| |||
Db    422 stdllnalkrvindpsykenvmklsriqhdqpvkpldravfwiefvmrhkgakhlrvaah 481

Qy    480 DLTWFQHYSIDVIGFLLTCVATAIFLFTKCFLFSCQKFNKTRK 522
          ||||||::|:|||||| |||| ||: ||| ||   || :  |
Db    482 dltwfqyhsldvigfllvcvatvifivtkccllfcfwkfarkak 524
```

RESULT 6
AAR26153
ID AAR26153 standard; Protein; 533 AA.
XX
AC AAR26153;
XX
DT 27-JAN-1993 (first entry)
XX
DE HUG-Br1.
XX
KW Bilirubin; UDP-glucuronosyltransferase; HUGBr1; HUGBr2;
KW monoglucuronide; diglucuronide.
XX
OS Homo sapiens.
XX

FH	Key	Location/Qualifiers
FT	Region	10..20
FT		/note= "putative membrane-insertion signal"
FT	Region	491..507
FT		/note= "putative membrane-anchoring peptide"
FT	Modified-site	102
FT		/note= "predicted Asn-linked glycosylation site"
FT	Modified-site	295
FT		/note= "predicted Asn-linked glycosylation site"
FT	Modified-site	347
FT		/note= "predicted Asn-linked glycosylation site"
FT	Misc-difference	158
FT		/note= "feature not labelled in specification"
FT	Misc-difference	181
FT		/note= "feature not labelled in specification"
FT	Misc-difference	228
FT		/note= "feature not labelled in specification"

XX
PN WO9212987-A.
XX
PD 06-AUG-1992.
XX
PF 10-JAN-1992; 92WO-US00282.
XX
PR 10-JAN-1991; 91US-0639453.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
XX
PI Owens IS, Ritter JK;
XX
DR WPI; 1992-284593/34.
DR N-PSDB; AAQ27369.
XX
PT Isolated gene locus UGT1, DNA segments and diagnostic probes -
PT for diagnosing Gilbert's disease and Crigler-Najjar syndrome
PT types I and II
XX
PS Disclosure; Fig 9A-I; 99pp; English.
XX
CC Two human liver bilirubin UDP-glucuronosyltransferase cDNAs have
CC been isolated. They are referred to as HUGBr1 (AAQ27369) and HUGBr2
CC (AAQ27370) (Ritter, et al., J. Biol. Chem. 266:1043-1047 (1991)) and,
CC upon expression individually in COS-1 cells, encode isoforms that
CC catalyse the formation of the two bilirubin monoglucuronides and
CC the diglucuronide.
CC The cDNAs contain identical 3' ends (1469 bp in length) to each
CC other and to that of the human phenol transferase cDNA, HLUGP1
CC (Harding et al., Proc. Natl. Aca. Sci. USA 85:8281 (1988)).
CC In contrast, they have unique 5' ends.
XX
SQ Sequence 533 AA;

Query Match 40.0%; Score 1120; DB 13; Length 533;
Best Local Similarity 45.0%; Pred. No. 3.8e-111;
Matches 232; Conservative 88; Mismatches 186; Indels 10; Gaps 6;

```
Qy     15 LFCV---GCGFCGKVLVWPCDMSHWLNVKVILEELIVRGHEVTVLTHSKPSLIDYRKPSA 71
          | ||        ||:|: | | ||||::   :::|  ||||: ||    |    | :  |
Db     16 llcvlgpvvshagkillipvdgshwlsmlgaiqqlqqrgheivvla---pdaslyirdga 72

Qy     72 L-KFEVVHMPQDRTEENEIFVDLALNVLPGLSTWQSVIKLNDFFVEIRGTLKMMCESFIY 130
              :   :|  | :  | || |  ||    |  | |||      :    |   |   ::
Db     73 fytlktypvpfqredvkesfvslghnvfendsflqrviktykkikkdsamllsgcshllh 132

Qy    131 NQTLMKKLQETNYDVMLIDPVIPCGDLMAELLAVPFVLTLRISVGGNMERSCGKLPAPLS 190
          |: ||  | |:::|||| || :||  ::|: |::| |  |  ::  ::|    : | | |
Db    133 nkelmaslaessfdvmltdpflpcspivaqylslptvfflh-alpcslefeatqcpnpfs 191

Qy    191 YVPVPMTGLTDRMTFLERVKNSMLSVLFHFWIQDYDYHFWEEFYSKALGRPTTLCETVGK 250
          ||| |::  :| ||||:|||| :::   :| : |  |  :    |: | |  |: : :
Db    192 yvprplsshsdhmtflqrvknmliafsqnf-lcdvvyspyatlaseflqrevtvqdllss 250

Qy    251 AEIWLIRTYWDFEFPQPYQPNFEFVGGLHCKPAKALPKEMENFVQSSGEDGIVVFSLGSL 310
          | :|| |: :  ::|:|  ||  ||||::|     | :| | :: :||| ||||||| |:
Db    251 asvwlfrsdfvkdyprpimpnmvfvgginclhqnplsqefeayinasgehgivvfslesm 310

Qy    311 FQNVTEEKANIIASALAQIPQKVLWRYKGKKPSTLGANTRLYDWIPQNDLLGHPKTKAFI 370
          : |:||  || || :||| |||| | :|| |  || |  |:|||||||| |:|||
Db    311 vseipekkamaiadalgkipqtvlwrytgtrpsnlanntilvkwlpqndllghpmtrafi 370

Qy    371 THGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKTMTSEDLLRALRTV 430
          || | :|:||:| :||||| :|:||||:||   |: ||| | :|  ||||||  | : |
Db    371 thagshgvyesicngvpmvmmplfgdqmdnakrmetkgagvtlnvlemtsedlenaqkav 430

Qy    431 ITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSAAHDLTWFQHYSID 490
          | | ||||| |||| :| |:||:||| ||||:||||||||| ||| |||||||:|::|:|
Db    431 indksykenimrlsslhkdrpvepldlavfwvefvmrhkgaphlrpaahdltwyqyhsld 490

Qy    491 VIGFLLTCVATAIFLFTKCFLFSCQK-FNKTRKIEK 525
          ||||||  : |  |:  ||  :  :|   |  :::|
Db    491 vigfllavlltvafitfkccaygyrkclgkkgrvkk 526
```

RESULT 7
AAR26154
ID AAR26154 standard; Protein; 534 AA.
XX
AC AAR26154;
XX
DT 27-JAN-1993 (first entry)
XX
DE HUG-Br2.
XX
KW Bilirubin; UDP-glucuronosyltransferase; HUGBr1; HUGBr2;
KW monoglucuronide; diglucuronide.
XX
OS Homo sapiens.

```
XX
FH   Key             Location/Qualifiers
FT   Region          12..22
FT                   /note= "putative membrane-insertion signal"
FT   Region          492..508
FT                   /note= "putative membrane-anchoring peptide"
FT   Modified-site   348
FT                   /note= "predicted Asn-linked glycosylation site"
FT   Misc-difference 282..285
FT                   /note= "residues encoded by TGCCAACGGGAAG !"
XX
PN   WO9212987-A.
XX
PD   06-AUG-1992.
XX
PF   10-JAN-1992;   92WO-US00282.
XX
PR   10-JAN-1991;   91US-0639453.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX
PI   Owens IS,  Ritter JK;
XX
DR   WPI; 1992-284593/34.
DR   N-PSDB; AAQ27369.
XX
PT   Isolated gene locus UGT1, DNA segments and diagnostic probes -
PT   for diagnosing Gilbert's disease and Crigler-Najjar syndrome
PT   types I and II
XX
PS   Disclosure; Fig 9A-I; 99pp; English.
XX
CC   Two human liver bilirubin UDP-glucuronosyltransferase cDNAs have
CC   been isolated.  They are referred to as HUGBr1 (AAQ27369) and HUGBr2
CC   (AAQ27370) (Ritter, et al., J. Biol. Chem. 266:1043-1047 (1991)) and,
CC   upon expression individually in COS-1 cells, encode isoforms that
CC   catalyse the formation of the two bilirubin monoglucuronides and
CC   the diglucuronide.
CC   The cDNAs contain identical 3' ends (1469 bp in length) to each
CC   other and to that of the human phenol transferase cDNA, HLUGP1
CC   (Harding et al., Proc. Natl. Aca. Sci. USA 85:8281 (1988)).
CC   In contrast, they have unique 5' ends.
XX
SQ   Sequence   534 AA;
```

```
  Query Match           37.8%;  Score 1058.5;  DB 13;  Length 534;
  Best Local Similarity 41.9%;  Pred. No. 1.5e-104;
  Matches  227;  Conservative   86;  Mismatches  172;  Indels   57;  Gaps    8;

Qy      11 LLLQLFCVGCGFCGKVLVWPCDMSHWLNVKVILEELIVRGHEVTVLT------HSKP--- 61
           ||| |        ||||| | | | ||:::  | ||  |||:  |||      | :
Db      16 lllllsvqpwaesgkvlvvptdgspwlsmrealrelhargheqavvltprgeyahqeekff 75

Qy      62 SLIDYRKPSALK------------FEVVHMPQDRTEENEIFVDLALNVLPGLSTWQSVI 108
           :|  |  |   |              ||  |: :  :    |  :::|
Db      76 tltayavpwtqkefdrvtlgytqgffetehilkrysrsmaimnnvsl------------ 122

Qy     109 KLNDFFVEIRGTLKMMCESFIYNQTLMKKLQETNYDVMLIDPVIPCGDLMAELLAVPFVL 168
                  |   |   ::|: |:: |  |::||:| |||  || ::|: |::| |
Db     123 ----------alhrccvellhnealirhlnatsfdvvltdpvnlcgavlakylsipavf 171

Qy     169 TLRISVGGNMERSCGKLPAPLSYVPVPMTGLTDRMTFLERVKNSM----LSVLFHFWIQD 224
             |  :  :::    : | | ||:|  :|  :| ||||:|||| :    || : | :
Db     172 fwry-ipcdldfkgtqcpnpssyipkllttnsdhmtflqrvknmlyplalsyichtfsap 230

Qy     225 YDYHFWEEFYSKALGRPTTLCETVGKAEIWLIRTYWDFEFPQPYQPNFEFVGGLHCKPAK 284
           |         |:   |  |: : :  | :|| |: :  ::|:|  ||  |:||::|   |
Db     231 y-----aslaselfqrevtvqdllssasvwlfrsdfvkdyprpimpnmvfiggincangk 285

Qy     285 ALPKEMENFVQSSGEDGIVVFSLGSLFQNVTEEKANIIASALAQIPQKVLWRYKGKKPST 344
           | :| | :: :||| ||||||| |:   : |:||  || || :||| |||| | :||
Db     286 plsqefeayinasgehgivvfslesmvseipekkamaiadalgkipqtvlwrytgtrpsn 345
```

```
Qy     345 LGANTRLYDWIPQNDLLGHPKTKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHM 404
           |  || |  |:||||||||| |:|||| | :|:||:| :|||| :|:||||:||   |
Db     346 lanntilvkwlpqndllghpmtrafithagshgvyesicngvpmvmmplfgdqmdnakrm 405

Qy     405 KAKGAAVEINFKTMTSEDLLRALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEF 464
           : ||| | :|   ||||||  | : || | ||||| |||| :| |:||:||| ||||:||
Db     406 etkgagvtlnvlemtsedlenaqkavindksykenimrlsslhkdrpvepldlavfwvef 465

Qy     465 VMRHKGAKHLRSAAHDLTWFQHYSIDVIGFLLTCVATAIFLFTKCFLFSCQK-FNKTRKI 523
           ||||||| ||| |||||||:|::|:|||||||  : |  |:  ||  :  :|   |  ::
Db     466 vmrhkgaphlrpaahdltwyqyhsldvigfllavlltvafitfkccaygyrkclgkkgrv 525

Qy     524 EK 525
           :|
Db     526 kk 527
```

```
RESULT   8
AAY57100
ID   AAY57100 standard; Protein; 245 AA.
XX
AC   AAY57100;
XX
DT   28-FEB-2000  (first entry)
XX
DE   UDP-glucuronosyltransferase 1 (UGT1) exons 2-5 amino acid sequence.
XX
KW   Uridine diphosphate-glucuronosyltransferase 1; UGT1; polymorphism; probe;
KW   glucuronic acid; Crigler-Najjar syndrome; Gilbert syndrome; jaundice;
KW   unconjugated hyperbilirubinaemia; drug metabolism; transgenic animal;
KW   pharmacogenetic screening; diagnose.
XX
OS   Homo sapiens.
XX
PN   WO9957322-A2.
XX
PD   11-NOV-1999.
XX
PF   04-MAY-1999;   99WO-US09702.
XX
PR   07-MAY-1998;   98US-0084807.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   Penny L,  Galvin M;
XX
DR   WPI; 2000-052981/04.
DR   N-PSDB; AAZ45118.
XX
PT   New nucleic acid representing polymorphisms in the human uridine
PT   diphosphate glucuronosyltransferase gene, used for diagnosis and
PT   evaluation of drug metabolism  -
XX
PS   Examples; Page 44-45; 63pp; English.
XX
CC   AAY57092-Y57100 are the amino acid sequences of exons 1A-1J of human
CC   uridine diphosphate-glucuronosyltransferase 1 (UGT1). The UGTs are a
CC   family of enzymes that catalyse the glucuronic acid conjugation of a
CC   wide range of endogenous and exogenous substrates including  phenols,
CC   alcohols, amines and fatty acids. Many of the reactions catalysed by
CC   UGTs result in toxic substances being converted to compounds which are
CC   more water soluble and are excreted. The invention relates to and
CC   identifies UGT1 polymorphisms (AAZ45004-Z45041). The polymorphism
CC   sequences are useful as probes for detecting UGT1 locus polymorphisms,
CC   indicative of altered UGT1 expression or activity. These polymorphisms
CC   are associated with Crigler-Najjar and Gilbert syndromes (unconjugated
CC   hyperbilirubinaemia) and drug metabolism. The genotyping of the UGT1 gene
CC   is used to predict the rate of metabolism of UGT1 substrates, possible
CC   drug-drug interactions and adverse side effects (i.e. to optimize drug
CC   dosage), and to screen for diseases caused by exposure to toxins and to
CC   study the effects of polymorphisms on enzymatic activity. The UGT1
CC   sequences, including polymorphisms, can also be used to produce the
```

```
CC   corresponding protein (or its fragments) or to generate transgenic
CC   animals or modified cells e.g. for pharmacogenetic screening.
XX
SQ   Sequence   245 AA;


  Query Match           27.0%;  Score 755.5;  DB 21;  Length 245;
  Best Local Similarity   59.7%;  Pred. No. 1.5e-72;
  Matches  142;  Conservative   34;  Mismatches   61;  Indels    1;  Gaps    1;

Qy     289 EMENFVQSSGEDGIVVFSLGSLFQNVTEEKANIIASALAQIPQKVLWRYKGKKPSTLGAN 348
           | | :: :||| |||||||||:   : |:||  || || :||| ||||| | :|| |  |
Db       1 efeayinasgehgivvfslgsmvseipekkamaiadalgkipqtvlwrytgtrpsnlann 60

Qy     349 TRLYDWIPQNDLLGHPKTKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHMKAKG 408
           | |  |:|||||||||| |:||||| | :|:||:| :||||| :|:||||:||   |: ||
Db      61 tilvkwlpqndllghpmtrafithagshgvyesicngvpmvmmplfgdqmdnakrmetkg 120

Qy     409 AAVEINFKTMTSEDLLRALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRH 468
           | | :|   ||||||  ||: || | ||||| |||| :| |:||:||| ||||:||||||
Db     121 agvtlnvlemtsedlenalkavindksykenimrlsslhkdrpvepldlavfwvefvmrh 180

Qy     469 KGAKHLRSAAHDLTWFQHYSIDVIGFLLTCVATAIFLFTKCFLFSCQK-FNKTRKIEK 525
           ||| ||| ||||||:|::|:|||||||  | |  |: ||  :  :|   |  :::|
Db     181 kgaphlrpaahdltwyqyhsldvigfllavvltvafitfkccaygyrkclgkkgrvkk 238


RESULT   9
AAB24025
ID   AAB24025 standard; Protein; 523 AA.
XX
AC   AAB24025;
XX
DT   25-JAN-2001  (first entry)
XX
DE   Human PRO1780 protein sequence SEQ ID NO:13.
XX
KW   Human; tumour; diagnosis; neoplastic disease; proliferation; cancer;
KW   identification; tumourigenesis; anticancer; detection.
XX
OS   Homo sapiens.
XX
PN   WO200053750-A1.
XX
PD   14-SEP-2000.
XX
PF   02-DEC-1999;   99WO-US28551.
XX
PR   08-MAR-1999;   99WO-US05028.
PR   01-SEP-1999;   99WO-US20111.
PR   29-OCT-1999;   99US-0162506.
PR   30-NOV-1999;   99WO-US28313.
PR   01-DEC-1999;   99WO-US28634.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Botstein D,  Goddard A,  Gurney AL,  Roy MA,  Watanabe CK,  Wood WI;
XX
DR   WPI; 2000-594320/56.
DR   N-PSDB; AAC58107.
XX
PT   Antibodies specific for PRO polypeptides, used to diagnose and inhibit
PT   the growth of tumors in mammals, and to identify inhibitors of PRO
PT   polypeptide activity or expression -
XX
PS   Claim 61; Fig 10; 226pp; English.
XX
CC   The present invention describes an antibody that binds to a human
CC   protein (I) selected from: PRO381; PRO1269; PRO1410; PRO1755; PRO1780;
CC   PRO3434; PRO1927; PRO3567; PRO1295; PRO1293; PRO1303; PRO4344; PRO4354;
CC   PRO4397; PRO4407; PRO1555; PRO1096; PRO2038; and PRO2262. (I) has
CC   anticancer activity and can be used to diagnose tumours in mammals, by
CC   detecting complex formation when the antibody is contacted with test
CC   cells. Increased expression of genes encoding (I) can also be detected
CC   to diagnose tumours. Agents which inhibit the activity of (I),
CC   especially the antibodies, or an antisense oligonucleotide which
CC   hybridises to genes encoding (I), can be used to inhibit tumour growth,
CC   preferably by inducing cell death. Methods from the present invention
CC   can be used to identify compounds which inhibit the biological activity
CC   of (I). AAC58019 to AAC58102 represent PCR primers and hybridisation
CC   probes used in examples from the present invention for human PRO
CC   sequences. AAC58103 to AAC58122 and AAB24021 to AAB24040 represent human
CC   PRO polynucleotide and protein sequences given in the exemplification of
CC   the present invention.
XX
SQ   Sequence   523 AA;


  Query Match           25.5%;  Score 714.5;  DB 21;  Length 523;
  Best Local Similarity   33.9%;  Pred. No. 1.3e-67;
  Matches  172;  Conservative   93;  Mismatches  208;  Indels   35;  Gaps   13;

Qy      34 SHWLNVKVILEELIVRGHEVTVLTHSK-PSLIDYRKPSALKFEVVH--MPQDRTEENEIF 90
           ||:| :  : : |   || ||:| | : | : |::|     ::|:    |:|   | :
Db      34 shyllmdrvsqilqdhghnvtmlnhkrgpfmpdfkkeek-syqviswlapedhqrefkks 92

Qy      91 VDLAL-NVLPGLSTWQSVIKLNDFFVEIRGTLKMMCESFIYNQTLMKKLQETNYDVMLID 149
             |  |   | |   ::::: : ::       | : |  |:  : :|  |:  |:|:::::
Db      93 fdffleetlggrgkfenllnvley-------lalqcshflnrkdimdslknenfdmvive 145

Qy     150 PVIPCGDLMAELLAVPFVLTLRISVGGNMERSCGKLPAPLSYVPVPMTGLTDRMTFLERV 209
             |  |:|| |  |||  |  |  |::|    || |||||||  : ||| | |  ||
Db     146 tfdycpfliaeklgkpfvailstsf-gslefg---lpiplsyvpvfrslltdhmdfwgrv 201

Qy     210 KNSMLSVLF----HFWIQDYDYHFWEEFYSKALGRPTTLCETVGKAEIWLIRTYWDFEFP 265
           || ::   |          :|    | |     |    |   : |||:| | : : |:|
Db     202 knflmffsfcrrqqhmqstfdntikehfte---gsrpvlshlllkaelwfinsdfafdfa 258

Qy     266 QPYQPNFEFVGGLHCKPAKALPKEMENFVQSSGEDGIVVFSLGSLF---QN--VTEEKAN 320
           :|  ||  :|||  || | :|:::|||:   |: | |: :|||:    ||  : :|  |
Db     259 rpllpntvyvgglmekpikpvpqdlenfiakfgdsgfvlvtlgsmvntcqnpeifkemnn 318

Qy     321 IIASALAQIPQKVLWRYK---GKKPSTLGANTRLYDWIPQNDLLGHPKTKAFITHGGMNG 377
             | | :|| |:|: :     |   | || :: ||:||:||| ||  : |:|||| |
Db     319 ----afahlpqgviwkcqcshwpkdvhlaanvkivdwlpqsdllahpsirlfvthggqns 374

Qy     378 IYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKTMTSEDLLRALRTVITDSSYK 437
           | ||| |||||||:|:|||| :|:  ::||   | |  | : :| |   :: :: |  ||
Db     375 imeaiqhgvpmvgiplfgdqpenmvrveakkfgvsiqlkklkaetlalkmkqimedkryk 434

Qy     438 ENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSAAHDLTWFQHYSIDVIGFLLT 497
             |:  | |    |: |  | | ||: |::  || ||:       | : |  ||  |||
Db     435 saavaasvilrshplsptqrlvgwidhvlqtggathlkpyvfqqpwheqylfdvfvfllg 494

Qy     498 CVATAIFLFTKCFLFSCQKFNKTRKIEK 525
             ::|  |    :       ||:::
Db     495 ltlgtlwlcgkllgmavwwlrgarkvke 522


RESULT  10
AAY99419
ID   AAY99419 standard; Protein; 523 AA.
XX
AC   AAY99419;
XX
DT   08-AUG-2000  (first entry)
XX
DE   Human PRO1780 (UNQ842) amino acid sequence SEQ ID NO:282.
XX
KW   Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
KW   transmembrane; secretion; immunoadhesion; pharmaceutical; screening.
XX
OS   Homo sapiens.
XX
PN   WO200012708-A2.
XX
```

PD 09-MAR-2000.
XX
PF 01-SEP-1999; 99WO-US20111.
XX
PR 01-SEP-1998; 98US-0098716.
PR 01-SEP-1998; 98US-0098749.
PR 01-SEP-1998; 98US-0098750.
PR 02-SEP-1998; 98US-0098803.
PR 02-SEP-1998; 98US-0098821.
PR 02-SEP-1998; 98US-0098843.
PR 09-SEP-1998; 98US-0099536.
PR 09-SEP-1998; 98US-0099596.
PR 09-SEP-1998; 98US-0099598.
PR 09-SEP-1998; 98US-0099602.
PR 09-SEP-1998; 98US-0099642.
PR 10-SEP-1998; 98US-0099741.
PR 10-SEP-1998; 98US-0099754.
PR 10-SEP-1998; 98US-0099763.
PR 10-SEP-1998; 98US-0099792.
PR 10-SEP-1998; 98US-0099808.
PR 10-SEP-1998; 98US-0099812.
PR 10-SEP-1998; 98US-0099815.
PR 10-SEP-1998; 98US-0099816.
PR 15-SEP-1998; 98US-0100385.
PR 15-SEP-1998; 98US-0100388.
PR 15-SEP-1998; 98US-0100390.
PR 16-SEP-1998; 98US-0100584.
PR 16-SEP-1998; 98US-0100627.
PR 16-SEP-1998; 98US-0100661.
PR 16-SEP-1998; 98US-0100662.
PR 16-SEP-1998; 98US-0100664.
PR 17-SEP-1998; 98US-0100683.
PR 17-SEP-1998; 98US-0100684.
PR 17-SEP-1998; 98US-0100710.
PR 17-SEP-1998; 98US-0100711.
PR 17-SEP-1998; 98US-0100919.
PR 17-SEP-1998; 98US-0100930.
PR 18-SEP-1998; 98US-0100848.
PR 18-SEP-1998; 98US-0100849.
PR 18-SEP-1998; 98US-0101014.
PR 18-SEP-1998; 98US-0101068.
PR 18-SEP-1998; 98US-0101071.
PR 22-SEP-1998; 98US-0101279.
PR 23-SEP-1998; 98US-0101471.
PR 23-SEP-1998; 98US-0101472.
PR 23-SEP-1998; 98US-0101474.
PR 23-SEP-1998; 98US-0101475.
PR 23-SEP-1998; 98US-0101476.
PR 23-SEP-1998; 98US-0101477.
PR 23-SEP-1998; 98US-0101479.
PR 24-SEP-1998; 98US-0101738.
PR 24-SEP-1998; 98US-0101741.
PR 24-SEP-1998; 98US-0101743.
PR 24-SEP-1998; 98US-0101915.
PR 24-SEP-1998; 98US-0101916.
PR 29-SEP-1998; 98US-0102207.
PR 29-SEP-1998; 98US-0102240.
PR 29-SEP-1998; 98US-0102307.
PR 29-SEP-1998; 98US-0102330.
PR 29-SEP-1998; 98US-0102331.
PR 30-SEP-1998; 98US-0102484.
PR 30-SEP-1998; 98US-0102487.
PR 30-SEP-1998; 98US-0102570.
PR 30-SEP-1998; 98US-0102571.
PR 01-OCT-1998; 98US-0102684.
PR 01-OCT-1998; 98US-0102687.
PR 02-OCT-1998; 98US-0102965.
PR 06-OCT-1998; 98US-0103258.
PR 06-OCT-1998; 98US-0103449.
PR 07-OCT-1998; 98US-0103314.
PR 07-OCT-1998; 98US-0103315.
PR 07-OCT-1998; 98US-0103328.
PR 07-OCT-1998; 98US-0103395.
PR 07-OCT-1998; 98US-0103396.
PR 07-OCT-1998; 98US-0103401.
PR 08-OCT-1998; 98US-0103633.
PR 08-OCT-1998; 98US-0103678.
PR 08-OCT-1998; 98US-0103679.
PR 08-OCT-1998; 98US-0103711.
PR 14-OCT-1998; 98US-0104257.
PR 20-OCT-1998; 98US-0104987.
PR 20-OCT-1998; 98US-0105000.
PR 20-OCT-1998; 98US-0105002.
PR 21-OCT-1998; 98US-0105104.
PR 22-OCT-1998; 98US-0105169.
PR 22-OCT-1998; 98US-0105266.
PR 26-OCT-1998; 98US-0105693.
PR 26-OCT-1998; 98US-0105694.
PR 27-OCT-1998; 98US-0105807.
PR 27-OCT-1998; 98US-0105881.
PR 27-OCT-1998; 98US-0105882.
PR 27-OCT-1998; 98US-0106062.
PR 28-OCT-1998; 98US-0106023.
PR 28-OCT-1998; 98US-0106029.
PR 28-OCT-1998; 98US-0106030.
PR 28-OCT-1998; 98US-0106032.
PR 28-OCT-1998; 98US-0106033.
PR 28-OCT-1998; 98US-0106178.
PR 29-OCT-1998; 98US-0106248.
PR 29-OCT-1998; 98US-0106384.
PR 29-OCT-1998; 98US-0108500.
PR 30-OCT-1998; 98US-0106464.
PR 03-NOV-1998; 98US-0106856.
PR 03-NOV-1998; 98US-0106902.
PR 03-NOV-1998; 98US-0106905.
PR 03-NOV-1998; 98US-0106919.
PR 03-NOV-1998; 98US-0106932.
PR 03-NOV-1998; 98US-0106934.
PR 10-NOV-1998; 98US-0107783.
PR 17-NOV-1998; 98US-0108775.
PR 17-NOV-1998; 98US-0108779.
PR 17-NOV-1998; 98US-0108787.
PR 17-NOV-1998; 98US-0108788.
PR 17-NOV-1998; 98US-0108801.
PR 17-NOV-1998; 98US-0108802.
PR 17-NOV-1998; 98US-0108806.
PR 17-NOV-1998; 98US-0108807.
PR 17-NOV-1998; 98US-0108867.
PR 17-NOV-1998; 98US-0108925.
PR 18-NOV-1998; 98US-0108848.
PR 18-NOV-1998; 98US-0108849.
PR 18-NOV-1998; 98US-0108850.
PR 18-NOV-1998; 98US-0108851.
PR 18-NOV-1998; 98US-0108852.
PR 18-NOV-1998; 98US-0108858.
PR 18-NOV-1998; 98US-0108904.
XX
PA (GETH) GENENTECH INC.
XX
PI Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;
XX
DR WPI; 2000-237871/20.
DR N-PSDB; AAA37101.
XX
PT New mammalian DNA sequences encoding transmembrane, receptor or
PT secreted PRO polypeptides, useful for screening of potential peptide or
PT small molecule inhibitors of the relevant receptor/ligand interactions
XX
PS Claim 12; Fig 160; 773pp; English.
XX
CC AAA37022 to AAA37144 encode the new isolated human transmembrane,
CC receptor or secreted PRO polypeptides given in AAY99340 to AAY99462. The
CC transmembrane and receptor PRO proteins can be used for screening of
CC potential peptide or small molecule inhibitors of the relevant
CC receptor/ligand interactions. The polypeptides and nucleotide sequences
CC encoding then have various industrial applications, including uses as

```
CC   pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent
CC   PCR primers and hybridisation probes used in the isolation of the PRO
CC   polypeptides from the present invention.
XX
SQ   Sequence   523 AA;


  Query Match             25.5%;  Score 714.5;  DB 21;  Length 523;
  Best Local Similarity   33.9%;  Pred. No. 1.3e-67;
  Matches  172;  Conservative   93;  Mismatches  208;  Indels   35;  Gaps   13;

Qy      34 SHWLNVKVILEELIVRGHEVTVLTHSK-PSLIDYRKPSALKFEVVH--MPQDRTEENEIF 90
           ||:| :  : : |   || ||:| | : | : |::|     ::|:    |:|   | :
Db      34 shyllmdrvsqilqdhghnvtmlnhkrgpfmpdfkkeek-syqviswlapedhqrefkks 92

Qy      91 VDLAL-NVLPGLSTWQSVIKLNDFFVEIRGTLKMMCESFIYNQTLMKKLQETNYDVMLID 149
             |  |   | |   ::::: : ::       | : |  |:  : :|  |:  |:|:::::
Db      93 fdffleetlggrgkfenllnvley-------lalqcshflnrkdimdslknenfdmvive 145

Qy     150 PVIPCGDLMAELLAVPFVLTLRISVGGNMERSCGKLPAPLSYVPVPMTGLTDRMTFLERV 209
             |  |:|| |  |||  |  |  |::|     || |||||||  : ||| | |  ||
Db     146 tfdycpfliaeklgkpfvailstsf-gslefg---lpiplsyvpvfrsllthmdfwgrv 201

Qy     210 KNSMLSVLF----HFWIQDYDYHFWEEFYSKALGRPTTLCETVGKAEIWLIRTYWDFEFP 265
           || ::   |          :|    | |     |    |   : |||:| | : : |:|
Db     202 knflmffsfcrrqqhmqstfdntikehfte---gsrpvlshlllkaelwfinsdfafdfa 258

Qy     266 QPYQPNFEFVGGLHCKPAKALPKEMENFVQSSGEDGIVVFSLGSLF---QN--VTEEKAN 320
           :|  ||  :|||| || | :|:::|||:   |: | |: :|||:    ||  : :|  |
Db     259 rpllpntvyvgglmekpikpvpqdlenfiakfgdsgfvlvtlgsmvntcqnpeifkemnn 318

Qy     321 IIASALAQIPQKVLWRYK---GKKPSTLGANTRLYDWIPQNDLLGHPKTKAFITHGGMNG 377
             | | :|| |:|: :     |   | || :: ||:||:||| ||  : |:|||| |
Db     319 ----afahlpqgviwkcqcshwpkdvhlaanvkivdwlpqsdllahpsirlfvthggqns 374

Qy     378 IYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKTMTSEDLLRALRTVITDSSYK 437
           | ||| |||||||:|:|||| :|:  ::||   | |  | : :| |   :: :: |  ||
Db     375 imeaiqhgvpmvgiplfgdqpenmvrveakkfgvsiqlkklkaetlalkmkqimedkryk 434

Qy     438 ENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSAAHDLTWFQHYSIDVIGFLLT 497
             |:  | |    |: |  | | ||: |::  || ||:       | : |  ||  |||
Db     435 saavaasvilrshplsptqrlvgwidhvlqtggathlkpyvfqqpwheqylfdvfvfllg 494

Qy     498 CVATAIFLFTKCFLFSCQKFNKTRKIEK 525
             ::|  |    :       ||:::
Db     495 ltlgtlwlcgkllgmavwwlrgarkvke 522


RESULT  11
AAB66168
ID   AAB66168 standard; protein; 523 AA.
XX
AC   AAB66168;
XX
DT   02-APR-2001  (first entry)
XX
DE   Protein of the invention #80.
XX
KW   Secreted; transmembrane; gene therapy.
XX
OS   Unidentified.
XX
PN   WO200078961-A1.
XX
PD   28-DEC-2000.
XX
PF   18-FEB-2000; 2000WO-US04342.
XX
PR   23-JUN-1999;   99US-0141037.
PR   20-JUL-1999;   99US-0144758.
PR   26-JUL-1999;   99US-0145698.
PR   01-SEP-1999;   99WO-US20111.
PR   29-OCT-1999;   99US-0162506.
PR   30-NOV-1999;   99WO-US28313.
PR   02-DEC-1999;   99WO-US28551.
PR   16-DEC-1999;   99WO-US30095.
PR   05-JAN-2000; 2000WO-US00219.
PR   06-JAN-2000; 2000WO-US00376.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker KP,  Botstein D,  Desnoyers L,  Eaton DL,  Ferrara N,  Fong S;
PI   Gao W,  Goddard A,  Godowski PJ,  Grimaldi CJ,  Gurney AL,  Hillan KJ;
PI   Pan J,  Paoni NF,  Roy MA,  Smith V,  Stewart TA,  Tumas D;
PI   Watanabe CK,  Williams PM,  Wood WI;
XX
DR   WPI; 2001-071395/08.
XX
PT   Secreted and transmembrane proteins and nucleic acids designated PRO,
PT   useful as hybridization probes, in chromosome and gene mapping and gene
PT   therapy -
XX
PS   Claim 1; Fig 160; 787pp; English.
XX
CC   The present invention relates to secreted and transmembrane proteins.
CC   These proteins and the DNA encoding them may be used as hybridization
CC   probes, in chromosome and gene mapping and in the generation of
CC   anti-sense RNA and DNA. They may also be used used to generate either
CC   transgenic animals or knockout animals which are in turn useful for
CC   development and screening of therapeutically useful reagents.
CC   The nucleic acids may also be used in gene therapy.
XX
SQ   Sequence   523 AA;


  Query Match             25.5%;  Score 714.5;  DB 22;  Length 523;
  Best Local Similarity   33.9%;  Pred. No. 1.3e-67;
  Matches  172;  Conservative   93;  Mismatches  208;  Indels   35;  Gaps   13;

Qy      34 SHWLNVKVILEELIVRGHEVTVLTHSK-PSLIDYRKPSALKFEVVH--MPQDRTEENEIF 90
           ||:| :  : : |   || ||:| | : | : |::|     ::|:    |:|   | :
Db      34 shyllmdrvsqilqdhghnvtmlnhkrgpfmpdfkkeek-syqviswlapedhqrefkks 92

Qy      91 VDLAL-NVLPGLSTWQSVIKLNDFFVEIRGTLKMMCESFIYNQTLMKKLQETNYDVMLID 149
             |  |   | |   ::::: : ::       | : |  |:  : :|  |:  |:|:::::
Db      93 fdffleetlggrgkfenllnvley-------lalqcshflnrkdimdslknenfdmvive 145

Qy     150 PVIPCGDLMAELLAVPFVLTLRISVGGNMERSCGKLPAPLSYVPVPMTGLTDRMTFLERV 209
             |  |:|| |  |||  |  |  |::|     || |||||||  : ||| | |  ||
Db     146 tfdycpfliaeklgkpfvailstsf-gslefg---lpiplsyvpvfrsllthmdfwgrv 201

Qy     210 KNSMLSVLF----HFWIQDYDYHFWEEFYSKALGRPTTLCETVGKAEIWLIRTYWDFEFP 265
           || ::   |          :|    | |     |    |   : |||:| | : : |:|
Db     202 knflmffsfcrrqqhmqstfdntikehfte---gsrpvlshlllkaelwfinsdfafdfa 258

Qy     266 QPYQPNFEFVGGLHCKPAKALPKEMENFVQSSGEDGIVVFSLGSLF---QN--VTEEKAN 320
           :|  ||  :|||| || | :|:::|||:   |: | |: :|||:    ||  : :|  |
Db     259 rpllpntvyvgglmekpikpvpqdlenfiakfgdsgfvlvtlgsmvntcqnpeifkemnn 318

Qy     321 IIASALAQIPQKVLWRYK---GKKPSTLGANTRLYDWIPQNDLLGHPKTKAFITHGGMNG 377
             | | :|| |:|: :     |   | || :: ||:||:||| ||  : |:|||| |
Db     319 ----afahlpqgviwkcqcshwpkdvhlaanvkivdwlpqsdllahpsirlfvthggqns 374

Qy     378 IYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKTMTSEDLLRALRTVITDSSYK 437
           | ||| |||||||:|:|||| :|:  ::||   | |  | : :| |   :: :: |  ||
Db     375 imeaiqhgvpmvgiplfgdqpenmvrveakkfgvsiqlkklkaetlalkmkqimedkryk 434

Qy     438 ENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSAAHDLTWFQHYSIDVIGFLLT 497
             |:  | |    |: |  | | ||: |::  || ||:       | : |  ||  |||
Db     435 saavaasvilrshplsptqrlvgwidhvlqtggathlkpyvfqqpwheqylfdvfvfllg 494

Qy     498 CVATAIFLFTKCFLFSCQKFNKTRKIEK 525
             ::|  |    :       ||:::
Db     495 ltlgtlwlcgkllgmavwwlrgarkvke 522
```

```
RESULT  12
AAB88348
ID   AAB88348 standard; Protein; 523 AA.
XX
AC   AAB88348;
XX
DT   23-MAY-2001  (first entry)
XX
DE   Human membrane or secretory protein clone PSEC0073.
XX
KW   Human; secretory protein; membrane protein; vaccine; gene therapy;
KW   rheumatoid arthritis; diabetes.
XX
OS   Homo sapiens.
XX
PN   EP1067182-A2.
XX
PD   10-JAN-2001.
XX
PF   07-JUL-2000; 2000EP-0114090.
XX
PR   08-JUL-1999;   99JP-0194179.
PR   11-JAN-2000; 2000JP-0118775.
PR   02-MAY-2000; 2000JP-0183766.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Kawai Y,  Sugiyama T,  Hayashi K;
XX
DR   WPI; 2001-093989/11.
DR   N-PSDB; AAF93775.
XX
PT   Nucleic acids encoding secretory proteins/membrane proteins, useful in
PT   gene therapy or as candidate target molecules in drug development -
XX
PS   Claim 1; SEQ ID 64; 609pp + CD ROM; English.
XX
CC   This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC   which encode human secretory or membrane proteins represented by
CC   AAB88317 - AAB88419. Included in the invention are primers
CC   AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
CC   cDNA sequences of the invention. The invention also includes methods for
CC   the production of antibodies directed against the proteins, and cDNA
CC   sequences, which can be used in vaccines. The polynucleotide sequences
CC   can be used in gene therapy. The polynucleotide sequences and the
CC   proteins they encode may be used in the prevention, treatment and
CC   diagnosis of diseases associated with inappropriate secretory
CC   protein/membrane protein expression. The nucleic acids and complementary
CC   sequences may also be used as DNA probes in diagnostic assays
CC   (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
CC   presence of similar nucleic acid sequences in samples. They may also be
CC   used to study the expression and function of secretory proteins/membrane
CC   polypeptides and their role in metabolism. The polypeptides may be used
CC   as antigens in the production of antibodies against them and in assays to
CC   identify modulators (agonists and antagonists) of expression and
CC   activity. The antibodies and antagonists may also be used as therapeutic
CC   agents to down regulate expression and activity. The antibodies may also
CC   be used as diagnostic agents for detecting the presence of the
CC   polypeptides in samples (e.g. by enzyme linked immunosorbant assay
CC   (ELISA). Examples of diseases which may be treated include rheumatoid
CC   arthritis and diabetes.
XX
SQ   Sequence   523 AA;


  Query Match             25.3%;  Score 708.5;  DB 22;  Length 523;
  Best Local Similarity   33.7%;  Pred. No. 5.6e-67;
  Matches  171;  Conservative   94;  Mismatches  208;  Indels   35;  Gaps   13;

Qy      34 SHWLNVKVILEELIVRGHEVTVLTHSK-PSLIDYRKPSALKFEVVH--MPQDRTEENEIF 90
           ||:| :  : : |   || ||:| | : | : |::|     ::|:    |:|   | :
Db      34 shyllmdrvsqilqdhghnvtmlnhkrgpfmpdfkkeek-syqviswlapedhqrefkks 92

Qy      91 VDLAL-NVLPGLSTWQSVIKLNDFFVEIRGTLKMMCESFIYNQTLMKKLQETNYDVMLID 149
             |  |   | |   ::::: : ::       | : |  |:  : :|  |:  |:|:::::
Db      93 fdffleetlggrgkfenllnvley-------lalqcshflnrkdimdslknenfdmvive 145

Qy     150 PVIPCGDLMAELLAVPFVLTLRISVGGNMERSCGKLPAPLSYVPVPMTGLTDRMTFLERV 209
             |  |:|| |  ||| |  |  |::|     || |||||||  : ||| | |  ||
Db     146 tfdycpfliaeklgkpfvailstsf-gslefg---lpiplsyvpvfrslltdhmdfwgrv 201

Qy     210 KNSMLSVLF----HFWIQDYDYHFWEEFYSKALGRPTTLCETVGKAEIWLIRTYWDFEFP 265
           || ::   |          :|    | |     |    |   : |||:| | : : |:|
Db     202 knflmffsfcrrqqhmqstfdntikehfte---gsrpvlshlllkaelwfinsdfafdfa 258

Qy     266 QPYQPNFEFVGGLHCKPAKALPKEMENFVQSSGEDGIVVFSLGSLF---QN--VTEEKAN 320
           :|  ||  :|||| || | :|:::|||:    : | |: :|||:    ||  : :|  |
Db     259 rpllpntvyvgglmekpikpvpqdlenfiakfedsgfvlvtlgsmvntcqnpeifkemnn 318

Qy     321 IIASALAQIPQKVLWRYK---GKKPSTLGANTRLYDWIPQNDLLGHPKTKAFITHGGMNG 377
             | | :|| |:|: :     |   | || :: ||:||:||| ||  : |:|||| |
Db     319 ----afahlpqgviwkcqcshwpkdvhlaanvkivdwlpqsdllahpsirlfvthggqns 374

Qy     378 IYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKTMTSEDLLRALRTVITDSSYK 437
           | ||| ||||||||:|:|:|||| :|:  ::||   | |  | : :| |   :: :: |  ||
Db     375 imeaiqhgvpmvgiplfgdqpenmvrveakkfgvsiqlkklkaetlalkmkqimedkryk 434

Qy     438 ENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSAAHDLTWFQHYSIDVIGFLLT 497
             |:  | |    |: |  | | ||: |::  || ||:       | : | :||  |||
Db     435 saavaasvilrshplsptqrlvgwidhvlqtggathlkpyvfqqpwheqylldvfvfllg 494

Qy     498 CVATAIFLFTKCFLFSCQKFNKTRKIEK 525
             ::|  |    :       ||:::
Db     495 ltlgtlwlcgkllgmavwwlrgarkvke 522


RESULT  13
AAG03280
ID   AAG03280 standard; Protein; 78 AA.
XX
AC   AAG03280;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein, SEQ ID NO: 7361.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   N-PSDB; AAC03286.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 13; SEQ ID 7361; 71pp + CD-ROM; English.
XX
CC   The present sequence is a polypeptide encoded by one of a large number
CC   of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC   were prepared from total human RNAs or polyA+ RNAs derived from 30
CC   different tissues. EST sequences usually correspond mainly to the 3'
```

```
CC   untranslated region (UTR) of the mRNA because they are often obtained
CC   from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC   isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC   those cases where longer cDNA sequences have been obtained, the full 5'
CC   UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC   ends and can therefore be used to obtain full length cDNAs and genomic
CC   DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC   chromosome mapping procedures. They are used to obtain upstream
CC   regulatory sequences and to design expression and secretion vectors.
XX
SQ   Sequence   78 AA;


  Query Match             14.5%;  Score 405;  DB 21;  Length 78;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-35;
  Matches   78;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     290 MENFVQSSGEDGIVVFSLGSLFQNVTEEKANIIASALAQIPQKVLWRYKGKKPSTLGANT 349
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 menfvqssgedgivvfslgslfqnvteekaniiasalaqipqkvlwrykgkkpstlgant 60

Qy     350 RLYDWIPQNDLLGHPKTK 367
           ||||||||||||||||||
Db      61 rlydwipqndllghpktk 78


RESULT  14
AAY29525
ID   AAY29525 standard; Protein; 129 AA.
XX
AC   AAY29525;
XX
DT   13-OCT-1999  (first entry)
XX
DE   Human lung tumour protein LT86-5 predicted amino acid sequence.
XX
KW   Human; lung tumour protein; therapy; diagnosis; lung cancer; vaccine;
KW   immunotherapy; detection; inhibition.
XX
OS   Homo sapiens.
XX
PN   WO9938973-A2.
XX
PD   05-AUG-1999.
XX
PF   26-JAN-1999;   99WO-US01642.
XX
PR   22-DEC-1998;   98US-0219245.
PR   28-JAN-1998;   98US-0015022.
PR   28-JAN-1998;   98US-0015029.
PR   18-MAR-1998;   98US-0040828.
PR   18-MAR-1998;   98US-0040831.
PR   23-JUL-1998;   98US-0122191.
PR   23-JUL-1998;   98US-0122192.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Frudakis TN,  Lodes MJ,  Mohamath R,  Reed SG;
XX
DR   WPI; 1999-479187/40.
DR   N-PSDB; AAZ07208.
XX
PT   Lung tumour specific polynucleotides for inhibiting the development
PT   of lung cancer
XX
PS   Example 2; Page 73; 171pp; English.
XX
CC   The present invention describes lung tumour specific polynucleotides
CC   and tumour antigens. AAZ07144 to AAZ07246 and AAZ08301 to AAZ08325
CC   represent specifically claimed polynucleotides, and AAY29486 to AAY29571
CC   represent amino acid sequences from the present invention. The lung
CC   tumour specific polynucleotides and polypeptides can be used in
CC   pharmaceutical compositions and vaccines to inhibit the development of
CC   lung cancer. They can also be used to detect lung cancer in a patient.
CC   Probes and antibodies derived from the lung tumour sequences are useful
CC   in detection of lung cancer.
XX
SQ   Sequence   129 AA;


  Query Match             14.0%;  Score 391;  DB 20;  Length 129;
  Best Local Similarity   58.1%;  Pred. No. 7.6e-34;
  Matches   75;  Conservative   18;  Mismatches   36;  Indels    0;  Gaps    0;

Qy     301 GIVVFSLGSLFQNVTEEKANIIASALAQIPQKVLWRYKGKKPSTLGANTRLYDWIPQNDL 360
           |||||||||:   : |:||  || || :||| ||||| | :|| |  || |  |:|||||
Db       1 givvfslgsmvseipekkavaiadalgkipqtvlwrytgtrpsnlanntilvqwlpqndl 60

Qy     361 LGHPKTKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKTMTS 420
           |||| |:|||||   :|: |:| :||||| :|:||||:||    : ||| | :|   |||
Db      61 lghpmtrafithasshgvnesicngvpmvmiplfgdqmdnakrretkgagvtlnvlemts 120

Qy     421 EDLLRALRT 429
           |||  ||::
Db     121 edledalks 129


RESULT  15
AAB44411
ID   AAB44411 standard; Protein; 129 AA.
XX
AC   AAB44411;
XX
DT   05-FEB-2001  (first entry)
XX
DE   Human lung tumour-specific antigen encoded by cDNA #21.
XX
KW   Lung tumour protein; lung cancer; cytostatic; vaccine.
XX
OS   Homo sapiens.
XX
PN   WO200060077-A2.
XX
PD   12-OCT-2000.
XX
PF   30-MAR-2000; 2000WO-US08560.
XX
PR   02-APR-1999;   99US-0285323.
PR   09-AUG-1999;   99US-0370838.
PR   30-DEC-1999;   99US-0476235.
PR   03-MAR-2000; 2000US-0518809.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Reed SG,  Lodes MJ,  Mohamath R,  Secrist H;
XX
DR   WPI; 2000-638466/61.
DR   N-PSDB; AAC79066.
XX
PT   Novel lung tumor polypeptides and polynucleotides, useful for
PT   detecting, monitoring or treating cancer, especially lung cancer -
XX
PS   Claim 1; Page 99; 243pp; English.
XX
CC   The present sequence is given in a specification relating to compounds
CC   for therapy and diagnosis of lung cancer. Polypeptides comprising at
CC   least an immunogenic part of a lung tumour protein are disclosed.
CC   The polypeptides are useful for inhibiting the development of cancer,
CC   especially lung cancer. Samples of T cells expressing the polypeptides
CC   may be used to inhibit the development of cancer. The polypeptides are
CC   also useful for detecting and monitoring the progression of cancer,
CC   especially lung cancer.
XX
SQ   Sequence   129 AA;
```

Query Match 14.0%; Score 391; DB 21; Length 129;
Best Local Similarity 58.1%; Pred. No. 7.6e-34;
Matches 75; Conservative 18; Mismatches 36; Indels 0; Gaps 0;

```
Qy     301 GIVVFSLGSLFQNVTEEKANIIASALAQIPQKVLWRYKGKKPSTLGANTRLYDWIPQNDL 360
           ||||||||||:   : |:||  || || :||| ||||| | :|| |  || |  |:|||||
Db       1 givvfslgsmvseipekkavaiadalgkipqtvlwrytgtrpsnlanntilvqwlpqndl 60

Qy     361 LGHPKTKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKTMTS 420
           |||| |:|||||   :|: |:| :||||| :|:||||:||    : ||| | :|   |||
Db      61 lghpmtrafithasshgvnesicngvpmvmiplfgdqmdnakrretkgagvtlnvlemts 120

Qy     421 EDLLRALRT 429
           |||  ||::
Db     121 edledalks 129
```

Search completed: August 13, 2001, 13:32:35
Job time: 4231 sec